# Structural Genomic Variation And Its Potential Role In Deer Speciation

Faezeh Azimi Chetabi[1], Aaron Shafer[1]

[1] 2140 East Bank Drive, Environmental and Life Sciences Graduate Program, Trent University, Peterborough, Ontario, Canada K9J 7B8

Correspondence: faezehazimichetabi@trentu.ca

**Abstract**

Speciation is a key driver of biodiversity and understanding its genomic underpinnings can be important for predicting and managing biodiversity. Structural variants (SVs) are large-scale (>50 bp) changes in the genome and have been implicated in adaptive divergence and reproductive isolation. We investigated the role of SVs in the speciation and divergence of two deer species (*Odocoileus* spp.) across their North American range. Using multiple long-read and a short-read datasets, our bioinformatics workflow revealed SVs and genomic features that were unique to each species. The majority of species-specific SVs were deletions and insertions suggesting that these variants may show higher likelihoods of fixation within populations. Further, while most SVs were intergenic, some genes were impacted, with 3 species-specific SVs showing signs of selection inferred from dN/dS. We also observed a reduced number of regulatory motifs found in fixed SVs compared to the rest of the genome. The SV-affected genes were often associated with reproduction and sensory adaptation, with such functions being relevant to fertility and deer biology and therefore providing insights into potential mechanisms leading to reproductive divergence.

**Keywords:** Comparative genomics, short-read, long-read, adaptation, genomic divergence, *Odocoileus*

## 1. Introduction

Speciation is a fundamental process driven most notably by ecological divergence and geographic isolation (Sobel, 2010), with the ultimate mode of divergence leaving tell-tale signals on the genome (Shang et al., 2023). Structural variants (SVs) have long been considered important in driving divergence, with early cytogenetic work pointing to chromosomal differences as barriers to reproduction (Rieseberg, 2001). In mammals for example, two distinct clades of Brocket deer (*Mazama* spp.) show karyotypic differences that contribute to reproductive isolation (Abril et al., 2010). Similarly, in donkey-horse hybrids (*Equus* spp.), failed chromosomal pairing during spermatogenic meiosis leads to sterility (Chandley et al, 1974; Chandley et al 1975). More recently, comprehensive reviews have highlighted that SVs, ranging from inversions to translocations, can influence speciation across diverse taxa (Wellenreuther et al., 2019; Berdan et al., 2023). The advent of high-throughput sequencing and novel bioinformatics tools has made it possible to assay SVs at base-pair resolution, opening a new line of inquiry into their role in adaptation and speciation across virtually all taxa.

While single nucleotide polymorphisms have historically dominated evolutionary genomics, SVs by definition affect larger stretches of DNA and can encompass entire genes and regulatory regions and may have stronger fitness consequences (Wellenreuther et al. 2019). Structural variants are generally considered features >50 bp and fall into several main types: deletions, duplications, inversions, insertions and translocations. Different sequencing and analytical techniques can be employed to detect SVs. Short-read (SR) data with paired read and split read analyses offer high-resolution breakpoint detection, though their effectiveness may depend on factors like insert size and coverage depth (Escaramís et al., 2015). Read depth analysis, often used with either SR or long-read (LR), provides information into SV presence by examining read density, but is less precise with breakpoint resolution (Escaramís et al., 2015). For LR data, de novo assembly and genome alignments are frequently employed, and offers the advantage of identifying novel SVs without relying on reference genomes, though it is more resource-intensive (Escaramís et al., 2015). Long-read sequencing is generally

preferable than SR sequencing for detecting complex and large SVs, as SR methods struggle with novel insertions and complex rearrangements (Mahmoud et al., 2019). More broadly, LR sequencing is proving transformative, revealing the prevalence and complexity of SVs across genomes and reshaping how we study divergence (Berdan et al., 2023).

Structural variants such as inversions and translocations can promote speciation through mechanisms like reproductive isolation (Homolka et al., 2007). Recombination suppression by SVs can also couple adaptive alleles and incompatibilities, reinforcing divergence even in the face of gene flow (Berdan et al., 2023; Wellenreuther et al., 2019). Autosomal translocations have been shown to cause male sterility in laboratory mice due to disrupted synapsis and meiotic silencing (Homolka et al., 2007), while Yang et al., (2024) showed convergent evolutionary signals of SVs associated with domestication, adaptation, and various traits. Further, Gompert et al. (2025) demonstrated that complex SVs, including inverted translocations, can independently arise in separate populations and repeatedly drive adaptive divergence by suppressing recombination and clustering functionally important genes, revealing a predictable mechanism of local adaptation and evolutionary change. Structural variants can also drive phenotypic diversity by influencing gene expression (Zhang et al, 2024), with enhancers often connecting genetic changes to species-specific phenotypes (Kaplow et al., 2023) and diseases (Fudenberg et al 2019; also referred to as enhancer-hijacking). Shi. et. al (2023) presented some evidence for enrichment of promoters and enhancers in specific SV categories. Collectively, these findings highlight SVs and enhancer regions as potential drivers of evolutionary divergence, shaping unique phenotypes and enabling evolution of traits across species.

The *Cervidae* (Deer) family offers an interesting case study for exploring SVs and genetic history. This family encompasses over 60 species with its origins dating back nearly 20 million years (Gonzalez and Duarte, 2020). The red brocket deer (*Mazama americana*) exhibits at least six distinct karyotypes across South America, suggesting the presence of multiple cryptic species (Cursino et al., 2014). Studies on hybrids between these karyotype lineages revealed post-zygotic reproductive isolation, with hybrids

showing sterility or subfertility due to chromosomal imbalances impacting ovarian structure, oocyte development, and embryonic genome activation (Cursino et al., 2014). Two closely related species, white-tailed deer (WTD; *Odocoileus virginianus*) and mule deer (MD; *O. hemionus*), have a relatively recent speciation history, appearing to have split in allopatry just over 1 million years ago (Ma) but also undergoing contemporary hybridization (Kessler et al., 2023). Kessler et al., (2023) suggested the existence of genetic incompatibilities given the lack of nuclear introgression and historical gene flow. Collectively, this raises the possibility that SVs might have played a role in deer speciation and genomic divergence.

In this study, we integrated multiple LR and SR sequencing datasets from MD and WTD across their geographic ranges to identify and characterize species-specific SVs. This comparative framework, which combines diverse sequencing approaches and detection tools (Figure 1), enables a comprehensive assessment of the evolutionary impact of SVs. We hypothesized that fixed SVs are important to deer speciation between the two deer species through two, non-mutually exclusive mechanisms: by driving adaptive divergence through changes in genic and regulatory regions, and by generating genomic incompatibilities that reduce hybrid fitness and reinforce reproductive isolation. By identifying consistent, species-specific SVs and examining their overlap with functional genomic regions and regulatory motifs, this study aims to better understand how SVs shapes adaptation, divergence, and the formation of species boundaries in *Odocoileus*.

## 2. Materials and Methods

### 2.1. Data generation and retrieval

We collated publicly available Illumina whole-genome short-read datasets from 79 deer (57 WTD and 22 MD; BioProject PRJNA830519); DNA was extracted from tissue samples (Qiagen DNeasy Blood and Tissue Kit) and sequenced at the Centre for Applied Genomics (Toronto, Canada) on an Illumina

HiSeqX platform (see Kessler et al., 2023 for more information). We also retrieved two LR datasets from NCBI (SRR6668252 for WTD and SRR15720448 for MD).

We generated new LR data from five deer semen samples (3 WTD and 2 MD); samples were first thawed on ice and cells were pelleted in an isotonic sperm wash solution to remove debris and reduce somatic cell contamination. DNA was extracted using the NEB Monarch® HMW DNA Extraction Kit for Cells & Blood (#T3050) UHMW protocol, with some modifications to the cell lysis step. Specifically, cells were digested at 56°C for 1 hour at 300 rpm with 100 µL Nuclei Prep Buffer, 100 µL Nuclei Lysis Buffer, 10 µL Proteinase K (NEB #P8107, 20 mg/ml), and 10 µL 1M DTT (GoldBio #DTT in dH2O, final concentration ~50 mM). An additional 20-minute digestion was then performed with 5 µL RNase A (NEB #T3018, 20 mg/ml). DNA quality control, CCS library preparation, and sequencing on the PacBio Revio sequencing instrument were then conducted. Collectively, there are four LR samples for WTD and three LR samples for MD. General characteristics and quality of the long-read datasets used in our analyses were assessed using SeqKit (Shen et al., 2016). For each LR sample, we summarized the total number of reads, and the minimum, average, and maximum read lengths.

We acquired two reference genome assemblies from the National Center for Biotechnology Information (NCBI) database: one of WTD (JAJQKH000000000) and the other of MD (JAJLRB000000000). To assess the completeness of these assemblies, we evaluated them using BUSCO (Simão et al, 2015), with the score of 98.3% for WTD and 98.0% for MD. We used the GENESPACE v1.1.4 R package (Lovell et al., 2022), that required OrthoFinder (Emms et al, 2019), DIAMOND (Buchfink et al, 2021) and MCScanX (Wang et al, 2012), to identify syntenic blocks and visualize macrosynteny relationships between WTD (*Odocoileus virginianus*), MD (*Odocoileus hemionus*), Caribou (*Rangifer tarandus*, CATKSN000000000), and domestic cattle (*Bos taurus*, NKLS00000000) (omitting *Cervus* for alignment compatibility reasons). To visualize chromosomal relationships, we generated riparian (macrosynteny) plots using the plot_riparian() function in GENESPACE. Color palettes and themes were customized using the R language. The WTD genome was used as the reference in the visualization.

## 2.2. SV detection pipeline

We employed two distinct analytical approaches to detect SVs (Figure 1). We used both LR datasets and SR data in conjunction with highest quality genome per species to detect species specific SVs (Figure S1-S2). Because reference genome assemblies can produce artefacts (Howe et al, 2021), and we recognized that there are multiple sequencing data and reference genome combinations, we identified the scenarios that generated species-specific SVs (Table 1). Specifically, the SVs detected in scenarios 4 and 5 are relatively unambiguous as species-specific SV. Likewise, scenarios 3 and 6 (both con-specific mapping) are likely only to arise via an assembly artefact or an SV singleton and are therefore not species-specific. Scenarios 1 and 2 are SVs detected with both reference assemblies, but not the sister species resequencing data; here we consider the SV to be real, and the SV is absent or misassembled in the reference. All SV indels are defined relative to the reference genome, meaning each reported indel represents reference-relative classifications rather than absolute evolutionary events.

For the SR pipeline, we removed adaptors from the raw fastq files using Trimmomatic (Bolger et al., 2014) with a MINLEN threshold of 36. We employed Bowtie2 (Langmead et al., 2012) to align all samples to both the reference genomes of WTD and MD. We used SAMtools (Danecek et al., 2021) to convert, sort, and index the SAM files. To ensure data integrity, duplicate reads were eliminated using Picard (Broad Institute, 2018) with REMOVE_SEQUENCING_DUPLICATES set to TRUE. We employed three different SV calling tools for SR data: Delly (Rausch et al., 2012), Manta (Chen et al., 2016), and Lumpy (Layer, 2014), all with default settings. Further refinement and comparative analysis of VCFs were performed using BCFtools (Danecek et al., 2021), retaining only structural variants with SUPP > 5, representing the number of sequencing reads supporting a given SV. We merged the SV call sets from Delly, Manta, and Lumpy using Jasmine (Kirsche et al, 2023), configured to confirm SVs found in at least two tools. Using BCFtools (Danecek et al., 2021), we filtered the structural variants to extract specific information for comparison, such as the number of SVs, their sizes in base pairs (bp) and zygosity status.

Our LR analysis pipeline followed a workflow similar to that described by Mérot et al. (2023). First, we mapped the LR datasets to the WTD and MD reference genomes using Winnowmap (Jain et al, 2020). The aligned SAM files were then converted to BAM format, sorted, and indexed with SAMtools (Danecek et al., 2021). Structural variants were identified using a combination of Sniffles (Sedlazeck et al., 2018), SVIM (Heller et al, 2019), and NanoVar (Tham et al, 2020). Finally, we employed Jasmine (Kirsche et al, 2023) to merge the SV calls into a single VCF file. We filtered the merged VCF files based on read support greater than 5.

Both LR and SR pipelines produced four mapping categories: MD2MD, MD2WTD, WTD2WTD, and WTD2MD (Figure S1-S2). For each individual, SVs were classified according to whether they were detected in: (1) both WTD2WTD and WTD2MD; (2) both MD2WTD and MD2MD; (3) only WTD2WTD; (4) only WTD2MD; (5) only MD2WTD; or (6) only MD2MD (Table 1). This classification framework enabled us to identify species-specific SVs (See Table 1, Figure S1-S2). The SVs from scenarios 3 and 6 were not included in further analysis,

To allow cross-referencing of features between species, we used MUMmer (Delcher et al., 2002) to align the WTD and MD reference genomes. Using the *show-coords* utility, we extracted the corresponding aligned intervals and identified regions shared between the two genomes and accurately lift over scaffold coordinates. To further refine our dataset and identify SVs with the potential relevance to speciation, we retained only those SVs present in 100% of the individuals within each scenario. In practice, this meant requiring SVs be detected in all 22 MD SR samples, or in all 57 WTD SR samples, or in all 4 WTD LR samples, or in all 3 MD LR samples, depending on the scenario. This filtering generated species-specific VCF datasets for both WTD and MD across the LR and SR data (Figure S1-S2). Finally, we quantified the overlap of these species-specific SVs between LR and SR analyses within each scenario.

### 2.3. Gene Impact, Selection Analysis and Enhancer Motifs

We annotated the MD and WTD genome for enhancer motifs using Homer (Heinz et al., 2010). The findMotifsGenome, annotatePeaks, and scanMotifGenomeWide tools were employed and filtered based on the average motif scores. We extracted the number of motifs detected and compared the number in intergenic regions, to those within 50 kb of any SV, and within 50 kb of species-specific SVs (up- and downstream); here, we normalized motif density, by dividing the number of motifs by the length of each region to produce a value per kilobase pair (kbp) and compared groups using a ANOVA. The resulting impact of SVs from both LR and SR analysis was annotated using SnpEff (Cingolani et al., 2012). To enable this, we created custom SnpEff databases using the genome FASTA files and their corresponding annotation GFF files. As with cross-referencing genomes, we transferred annotations between assemblies using Mummer (Delcher et al., 2002) and custom Python scripts.

Lastly, we calculated the ratio of nonsynonymous to synonymous substitution rates (dN/dS) for specific genes overlapping and the fixed (high confidence) SV categories; we prepared coding sequences (CDS) and corresponding protein sequences for the same species used for synteny analysis (omitting *Alces* for alignment compatibility reasons). Single copy orthologous genes were identified using OrthoFinder (Emms et al, 2019) with DIAMOND (Buchfink et al, 2021) for sequence similarity searches, MAFFT (Katoh et al., 2005) for multiple sequence alignment, and FastTree (Price etal, 2009) for species tree inference. For each orthogroup, corresponding CDS sequences were extracted using seqtk (Li, 2012) and aligned at the codon level using PAL2NAL (Suyama et al, 2006). Gene sequences were used as input for BUSTED (Murrell et al, 2025) method implemented in HyPHy (Kosakovsky Pond et al., 2005) to estimate dN/dS ratios. The predefined species tree topology (Cow, (Elk, (Caribou, ((MuleDeer, WhiteTailedDeer))))); was used in all analyses. Evidence of selection on individual MD or WTD lineages using the foreground analysis was assessed using likelihood ratio tests with significance determined at $p < 0.05$ after Bonferroni-Holm correction.

## 3. Results

**3.1.** Genome SV exploration

The Cervidae genomes appear to be largely syntenic at a macro-level (Figure 2). Average sequencing coverage for the SR datasets was approximately 6X; the corresponding LR datasets showed variation in coverage, but all were > 7X (Table 2). Structural variant distribution across all individuals in the SR and LR merged SV datasets (i.e. 2/3 methods detected the SV) are shown in Figure 3 for mapping to the conspecific genome (i.e. WTD2WTD and MD2MD). Here, the median SV size in bp presented WTD then MD for deletion was 206 and 239, duplication was 1817 and 2027, insertion was 84 and 89 bps, inversion was 1294 and 2501 bps and translocation was 610 and 850; average individual SV homozygosity was >50% for indels (Table S1). In the LR merged SV dataset, the median SV size in bp presented WTD then MD for deletion was 136 and 133, duplication was 9821 and 5531, insertion was 109 and 132 bps, inversion was 570 and 1178 bps and translocation was 201 and 746; average individual SV homozygosity was 30%-70% for indels (Table S1).

### **3.2.** Species specific SVs, Genome and Enhancer Impact

We next focused on species-specific SVs that were present in 100% of samples within each species in Scenarios 1,2,4,5 (Table 1). In total we identified 56 unique SVs in 100% of SR samples (Table 3). In contrast, we identified 123,384 unique SVs in 100% of LR samples (Table 4). The majority of predicted gene-impacts were intronic (Table 3 and 4). We assessed the overlap between SR- and LR-derived SVs by considering only those variants located within exonic, intronic, or gene-level regions (Table 5); intergenic SVs were excluded from this comparison. Of note 3 of genes had significant dN/dS values in WTD and MD; we identified notable genes in fixed SVs as it pertains to deer biology and potential reproductive isolation (Table S2). For both MD and WTD, enhancer motifs were elevated in intergenic regions, with those adjacent to fixed SVs having the lowest density (Table S3; ANOVA $F = 45.25$, $p < 0.01$).

## 4. Discussion

The genomes of recent and hybridizing species provide a window into the mechanisms that drive biodiversity, adaptation, and reproductive isolation. We compared both SR and LR data on two closely related and hybridizing deer species, the white-tailed deer and mule deer, to identify SVs fixed between species. By integrating data from multiple LR and SR sequencing, we identified SVs within and between these species, their predicted impact on genes, and general relationship to regulatory motifs. Genome analyses has detected negligible historical introgression between these species, despite contemporary hybridization (Kessler et al., 2023), suggesting that much of the observed genomic divergence reflects drift and isolation rather than ongoing exchange. This lack of gene flow complicates the identification of barrier loci, since patterns of divergence cannot be readily distinguished from neutral processes, but there is very little evidence for any meaningful nuclear gene-flow (Kessler et al. 2023). While the presence of species-specific SVs reflects major evolutionary events, we acknowledge that fixed differences alone cannot demonstrate a causal role in reproductive isolation, as divergence also accumulates neutrally in the absence of gene flow. Nonetheless, the fixed SVs that encompass genes and regulatory regions provides testable hypotheses and clues to species divergence and reproductive isolation.

**General characteristics of SVs in *Odocoileus***

Our analysis revealed a predominance of insertions and deletions among SVs in deer. This finding is consistent with observations reported in other studies (Shi et al., 2023; Zhang et al., 2023), notably, Shi et al. (2023) identified deletions as the most frequent SV type in human populations. This pattern could be attributed to various mechanisms, including replication errors or unequal homologous recombination (Wold et al., 2023). Berdan et al., (2023) summarized how SVs, particularly when heterozygous, can suppress recombination, but homozygous deletions and insertions, by their nature, do not inherently disrupt the linear arrangement of genetic material, minimizing their impact on recombination. Duplications and translocations, particularly in their heterozygous form, can result in complex genomic rearrangements that disrupt gene order and heighten the risk of recombination errors. Zhang et al. (2023)

discussed how gene duplications can drive speciation by disrupting gene function or fostering novel interactions, with empirical evidence showing duplications clearly underlying incompatibilities (Zuellig et al 2018; Bikard et al 2009). While the species-specific duplications were among the least common SV we detected (Table 3 and 4), these are worth further functional exploration, especially those containing genic regions.

Most of the SV impacts we observed were intergenic, and similar trends were reported by Yang et al. (2024) and David et al. (2024). These studies, like ours, demonstrated that a significant portion of SVs localized to intergenic regions in sheep, goats, and wild bird populations. This dominance of intergenic impacts can be attributed to the vast amount of non-coding DNA within the genome; however, these regions also contain important regulatory elements, which comprise a significant portion of the mammalian genome (Villar et al, 2015). David et al. (2024) observed that a high percentage of deletions and inversions in intergenic regions were classified as modifiers, suggesting they could have phenotypic effects. Our findings provide a similar perspective where 80.11% of deletions and 72.99% of inversions occurred in intergenic regions and 18.44% of intergenic deletions and 15.99% of intergenic inversions were annotated as modifiers, reiterating the potential consequential impacts SVs can have outside genic regions. The reduced density of enhancers in SVs, particularly those in fixed SVs (Table S3), highlight the influence of intergenic space in the genome when it comes to promoters and enhancers (ENCODE Project Consortium, 2012). Shi et al. (2023) found a genome-wide positive correlation between enhancer and promotors and SV density. The reduction in motifs in the fixed SVs is an interesting observation in deer; while enhancer and silencer action are often cell dependent (Panigrahi et al, 2021), this pattern is consistent with a positional or cis-ruption effect (see Kleinjan and Coutinho, 2009), meaning the regulatory architecture is impacted, with the molecular and phenotypic impact in deer warranting further investigation.

**Species-species SVs, adaptation and potential incompatibilities**

The identification of fixed SVs highlights potential connections to adaptive divergence. Recent studies in Eucalyptus species (Ferguson et al., 2024) showed the role of shared SVs in adaptation and speciation with significant number of SVs being enriched for genes involved in adaptive traits. Likewise, work on stick-insects showed clustering functionally important genes (Gompert et al. 2025). In deer, patterns were nuanced but did align with predictions and previous work, with deletions being the most common fixed SV. This can be functionally important as Derks et al. (2018) showed in pigs that a large homozygous deletion that exhibited the highest penetrance. In slight contrast, Redin et al. (2017) found that inversions and translocations van have stronger association congenital abnormalities. Accordingly, the higher rate fixed deletions between species might point to a role in adaptive divergence and speciation.

Several presumed gene impacts were identified that align with known aspects of deer biology and presumed speciation (Kessler et al., 2024). Notably, there were multiple genes with known links to male gonad development and sperm production (Table S2), which is important because hybrid deer show evidence of reduced sperm function (Derr et al., 1991). Similarly, both Bracewell et al. and Dowle et al. (2017) showed how SVs on neo-sex chromosomes drive hybrid male sterility, demonstrating that chromosomal rearrangements can underpin the genetic basis of speciation via genes directly involved in fertility. We also observed SVs impacting genes linked to olfactory function (Table S2), similar processes were identified in the SNP based approach of Kessler et al. (2024). Differences in olfactory genes might facilitate ecological partitioning between WTD and MD, contributing to their speciation. The role of olfactory genes in mate recognition and behavioral adaptation has been seen in house mice (North et al., 2020), where copy number variations in olfactory receptor genes influenced mate choice and reinforced reproductive isolation. This is particularly relevant in deer, as differences in olfactory responses might not only dictate ecological adaptations but also impact mate recognition, creating a dual mechanism for speciation through both environmental and sexual selection. Collectively, such fixed SVs

and the corresponding phenotypic variation might implicate possible Dobzhansky-Muller incompatibilities in maintaining species boundaries in deer.

Further, several genes associated with species-specific SVs exhibited evidence of positive selection in both MD and WTD. The SV impacting genes with positive dN\dS values (3 total), were translocations and duplications, but less clear links to reproductive or ecological divergence (Table S2). Here we note one limitation in evaluating the evolutionary consequences of duplications lies in SV detection, as current SR-based tools identify but do not reconstruct the SV, with the absence of the SV in the assembly reflective of the common practice of purging duplications (Guan et al, 2020). As a result, dN/dS calculations are confined to the sequence of the initial (homologous) duplicated region, though LR approaches will allow for assess the evolution of duplicated regions.

**Differences between SV detection methods and future directions**

The accurate estimation of SV size can be influenced by the tools and methods employed (Sedlazeck et al., 2018; Jeffares et al., 2017). The accuracy and sensitivity of SV callers vary significantly, as some tools are better suited for detecting larger SVs, while others excel at identifying smaller changes (Sedlazeck et al., 2018; Sudmant et al., 2015; Jeffares et al., 2017). This inherent bias across tools shapes the type of SV detected and the observed size distribution. To account for this, we employed two distinct data streams for SV detection to evaluate genome differences between MD and WTD, both reflective of commonly available datasets. Our comparison of SR and LR SV datasets showed differences that shape how structural variation is represented in genomic studies. As demonstrated in earlier benchmarking and evolutionary studies (Pei et al., 2024; Mahmoud et al., 2019), LR technologies detect substantially more SVs across nearly all classes, and our results reflect this same pattern (Tables 3 and 4). This discovery is expected, as long reads traverse repetitive and low-complexity regions where most SV breakpoints occur, allowing callers to resolve thousands of smaller and intermediate events that SR platforms fail to detect. In contrast, the SR datasets captured markedly fewer SVs across all classes and exhibited patterns consistent with well-known short-read constraints (Mérot et al, 2023). The large

SR sample size and requirement of 100% SV detection in all samples (both LR and SR), might be overly stringent, but provided confidence in the observed calls, especially those overlapping between data sets (Table 5). While the absolute counts between SR and LR calls differed (Table 3 and 4), any non-overlapping SV can be presumed a false-positive or false-negative, mostly likely occurring in the SR data. The varied scenarios (Table 1) also allowed for assembly artefacts and singletons to be removed, where the only explanation for SVs detected only in either WTD2WTD or MD2MD comparison (scenarios 1 and 2) is that assembly differs from the conspecific resequencing data. Ultimately, the combined tally of SVs (Table 5), reflect high-confidence calls, and are likely evolutionarily relevant.

Future research can build on our findings through several approaches aimed at confirming the role of SVs in speciation. Mouse models would offer experimental approaches to validate, but analyzing known hybrids and backcrosses in deer should allow assessing the distribution and inheritance patterns of SVs. RNA-seq combined with chromatin accessibility assays could allow us to measure how SVs and the enhancer-rich regions alter gene regulation in both species. Lastly, expanding the scope to comparative analyses across other cervid species will further clarify whether similar classes of SVs repeatedly contributed to divergence and allow for reconstructing the ancestral state of species-specific SVs in *Odocoileus*.

**Acknowledgements:** We thank Alan Cain, Steeve Côté, Anh Dao, Marco Festa-Bianchet, Brad Fulk, Eric Hoffman, Levi Jaster, Daniel Koelsch, Emily Latch, Brent Patterson, Joe Nocera, Don Stewart, Russell Easy, Jonathan Shaw, David Walter, and Jon Wheeler for providing samples. We thank two reviewers for providing valuable feedback.

**Data availability:** All SR and some of LR sequencing samples are publicly available in the NCBI database under BioProject accession PRJNA830519. New LR data has been deposited (PRJA ID pending). And ortholog alignments in Figshare (https://doi.org/10.6084/m9.figshare.31830535). The

code supporting this work is available in our lab's GitLab repository: https://gitlab.com/WiDGeT_TrentU/graduate_theses/-/tree/master/AzimiChetabi/Chapter%20one

**Author Contribution:** We had no external contributions.

**Funding:** We were supported by Trent University's International Graduate Scholarship, an Ontario Graduate Scholarship, and Dean's PhD Scholarship. This work was supported by the Natural Sciences and Engineering Research Council of Canada Discovery Grant (grant number: RGPIN-2017-03934), ComputeCanada Resources for Research Groups (grant number: RRG gme-665-ab), Canada Foundation for Innovation: John R. Evans Leaders Fund, and Ontario Early Researcher Award (grant number: #36905).

**Conflict of interest statement:** The authors declare no competing interests.

## References


Abril, V. V., E. A. G. Carnelossi, S. González, and J. M. B. Duarte. 2010. Elucidating the evolution of the red brocket deer Mazama americana complex (Artiodactyla; Cervidae). Cytogenetic and Genome Research 128:177–187.

Berdan, E. L., T. G. Aubier, S. Cozzolino, R. Faria, J. L. Feder, M. D. Giménez, M. Joron, J. B. Searle, and C. Mérot. 2024. Structural Variants and Speciation: Multiple Processes at Play. Cold Spring Harb Perspect Biol. 16:a041446.

Bikard, D., D. Patel, C. Le Metté, V. Giorgi, C. Camilleri, M. J. Bennett, and O. Loudet. 2009. Divergent evolution of duplicate genes leads to genetic incompatibilities within Arabidopsis thaliana. Science 323:623–626.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1. PMID: 24695404; PMCID: PMC4103590.

Buchfink B, Reuter K, Drost HG. (2021). Sensitive protein alignments at tree-of-life scale using DIAMOND. Nature Methods, 18, 366-368.

Broad Institute. 2018. Picard Toolkit [GitHub repository]. Retrieved from http://broadinstitute.github.io/picard/

Bracewell, R. R., B. J. Bentz, B. T. Sullivan, and J. M. Good. 2017. Rapid Neo-Sex Chromosome Evolution and Incipient Speciation in a Major Forest Pest. Nat. Commun. 8:1593.

Chandley, A. C., R. C. Jones, H. M. Dott, W. R. Allen, and R. V. Short. 1974. Meiosis in interspecific equine hybrids. I. The male mule (Equus asinus × E. caballus) and hinny (E. caballus × E. asinus). Cytogenetics and Cell Genetics 13:330–341.

Chandley, A. C., R. V. Short, and W. R. Allen. 1975. Cytogenetic studies of three equine hybrids. Journal of Reproduction and Fertility Supplement 23:356–370.

Chen, X., O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger, M. Källberg, et al. 2016. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32:1220–1222.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6:80–92.

Cursino, M. S., M. B. Salviano, V. V. Abril, et al. 2014. The role of chromosome variation in the speciation of the red brocket deer complex: The study of reproductive isolation in females. BMC Evolutionary Biology 14:40.

Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li. 2021. Twelve years of SAMtools and BCFtools. GigaScience 10:giab008.

David, G., A. Bertolotti, R. Layer, D. Scofield, A. Hayward, T. Baril, H. A. Burnett, E. Gudmunds, H. Jensen, and A. Husby. 2024. Calling structural variants with confidence from short-read data in wild bird populations. Genome Biology and Evolution 16:evae049.

Delcher, A. L., A. Phillippy, J. Carlton, and S. L. Salzberg. 2002. Fast algorithms for large-scale genome alignment and comparison. Nucleic Acids Research 30:2478–2483.

Derks, M. F. L., M. S. Lopes, M. Bosse, O. Madsen, B. Dibbits, B. Harlizius, et al. 2018. Balancing selection on a recessive lethal deletion with pleiotropic effects on two neighboring genes in the porcine genome. PLoS Genetics 14:e1007661.

Derr, J. N., D. W. Hale, D. L. Ellsworth, and J. W. Bickham. 1991. Fertility in an F1 male hybrid of white-tailed deer (*Odocoileus virginianus*) x mule deer (*O. hemionus*). J Reprod Fertil. 93:111-7.

Dowle, E. J., R. R. Bracewell, M. E. Pfrender, K. E. Mock, B. J. Bentz, and G. J. Ragland. 2017. Reproductive Isolation and Environmental Adaptation Shape the Phylogeography of Mountain Pine Beetle (Dendroctonus Ponderosae). Mol. Ecol. 26:6071–6084.

Emms D.M., Kelly S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology, 20:238.

Escaramís, G., E. Docampo, and R. Rabionet. 2015. A decade of structural variants: Description, history, and methods to detect structural variation. Briefings in Functional Genomics 14:367–379.

Ferguson, S., A. Jones, K. Murray, R. L. Andrew, B. Schwessinger, H. Bothwell, J. Borevitz, 2024. Exploring the role of polymorphic interspecies structural variants in reproductive isolation and adaptive divergence in Eucalyptus. GigaScience 13:giae029.

Fudenberg, G., and K. S. Pollard. 2019. Chromatin features constrain structural variation across evolutionary timescales. Proc Natl Acad Sci U S A 116:2175–2180.

Gompert, Z., J. L. Feder, T. L. Parchman, et al. 2025. Adaptation repeatedly uses complex structural genomic variation. *Science* 388(6744):eadp3745.

Gonzalez, S., and J. M. B. Duarte. 2020. Speciation, evolutionary history and conservation trends of Neotropical deer. *Mastozoología Neotropical* 27(SI):35–46.

Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 2020 May 1;36(9):2896-2898. doi: 10.1093/bioinformatics/btaa025. PMID: 31971576; PMCID: PMC7203741.

Heller, D., & Vingron, M. (2019). SVIM: Structural variant identification using mapped long reads. *Bioinformatics*, *35*(17), 2907–2915.

Heinz, S., C. Benner, N. Spann, E. Bertolino et al. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell 38:576-589.

Homolka, D., R. Ivánek, J. Capkova, P. Jansa, and J. Forejt. 2007. Chromosomal rearrangement interferes with meiotic X chromosome inactivation. Genome Research 17:1431–1437

Howe, K., W. Chow, J. Collins, S. Pelan, D.-L. Pointon, Y. Sims, J. Torrance, A. Tracey, and J. Wood. 2021. Significantly improving the quality of genome assemblies through curation. *GigaScience* 10:1–9.

Jain, C., Rhie, A., Zhang, H., Chu, C., Walenz, B., Koren, S., & Phillippy, A. (2020). Weighted minimizer sampling improves long read mapping. Bioinformatics, 36(Suppl_1), i111–i118.

Jeffares, D. C., C. Jolly, M. Hoti, D. Speed, L. Shaw, C. Rallis, et al. 2017. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. Nature Communications 8:14061.

Kaplow, I. M., M. T. Maurano, A. Kundaje, et al. 2023. Relating enhancer genetic variation across mammals to complex phenotypes using machine learning. *Science* 380(6648):eabm7993.

Katoh, K., and D. M. Standley. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30(4):772–780.

Kessler C, Shafer ABA. Genomic analyses capture the human-induced demographic collapse and recovery in a wide-ranging cervid. Mol Biol Evol. 2024; 41: msae038.

Kessler, C., E. Wootton, and A. B. A. Shafer. 2023. Speciation without gene-flow in hybridizing deer. Mol Ecol. 32:1117-1132.

Kirsche M, Prabhu G, Sherman R, Ni B, Battle A, Aganezov S, Schatz MC. Jasmine and Iris: population-scale structural variant comparison and analysis. Nat Methods. 2023 Mar;20(3):408-417. Epub 2023 Jan 19. PMID: 36658279; PMCID: PMC10006329.

Kleinjan, D.-J., and P. Coutinho. 2009. Cis-ruption mechanisms: disruption of cis-regulatory control as a cause of human genetic disease. Brief Funct Genomics Proteomics 8:317–332.

Kosakovsky Pond, S. L., S. D. W. Frost, and S. V. Muse. 2005. HyPhy: Hypothesis testing using phylogenies. *Bioinformatics* 21(5):676–679.

Langmead, B., and S. L. Salzberg. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.

Layer, R. M., C. Chiang, A. R. Quinlan, and I. M. Hall. 2014. LUMPY: A probabilistic framework for structural variant discovery. Genome Biology 15:R84.

Lovell, J. T., Sreedasyam, A., Schranz, M. E., Wilson, M., Carlson, J. W., Harkess, A., Emms, D., Goodstein, D. M., & Schmutz, J. (2022). GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. *eLife*, 11, e78526.

Li H. 2012. Seqtk: Toolkit for processing sequences in FASTA/Q formats. https://github.com/lh3/seqtk

Mahmoud, M., N. Gobet, D. I. Cruz-Dávalos, et al. 2019. Structural variant calling: the long and the short of it. Genome Biology 20:246.

Mérot C, Stenløkk KSR, Venney C, Laporte M, Moser M, Normandeau E, Árnyasi M, Kent M, Rougeux C, Flynn JM, Lien S, Bernatchez L. 2023, Genome assembly, structural variants, and genetic differentiation between lake whitefish young species pairs (Coregonus sp.) with long and short reads. Mol Ecol; 32: 1458–1477.

Murrell, B., S. Weaver, M. D. Smith, J. O. Wertheim, S. Murrell, A. Aylward, K. Eren, T. Pollner, D. P. Martin, D. M. Smith, K. Scheffler, and S. L. Kosakovsky Pond. 2015. Gene-wide identification of episodic selection. *Molecular Biology and Evolution* 32:1365–1371.

North, H. L., P. Caminade, D. Severac, K. Belkhir, and C. M. Smadja. 2020. The role of copy-number variation in the reinforcement of sexual isolation between the two European subspecies of the house mouse. Phil. Trans. R. Soc. B 375:20190540.

Panigrahi, A., and B. W. O'Malley. 2021. Mechanisms of enhancer action: the known and the unknown. Genome Biol 22:108.

Pei, Y., Tanguy, M., Giess, A., Dixit, A., Wilson, L. C., Gibbons, R. J., Twigg, S. R. F., Elgar, G., & Wilkie, A. O. M. (2024). A comparison of structural variant calling from short-read and nanopore-based whole-genome sequencing using optical genome mapping as a benchmark. *Genes*, 15(7), 925.

Price MN, Dehal PS, Arkin AP. 2009. FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. Molecular Biology and Evolution 26:1641–1650.

Rausch, T., T. Zichner, A. Schlattl, A. M. Stutz, V. Benes, and J. O. Korbel. 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics 28:i333–i339.

Redin, C., H. Brand, R. L. Collins, T. Kammin, E. Mitchell, J. C. Hodge, C. Hanscom, V. Pillalamarri, C. M. Seabra, M. A. Abbott, O. A. Abdul-Rahman, E. Aberg, R. Adley, S. L. Alcaraz-Estrada, F. S. Alkuraya, Y. An, M. A. Anderson, C. Antolik, K. Anyane-Yeboa, J. F. Atkin, T. Bartell, J. A. Bernstein, E. Beyer, I. Blumenthal, E. M. Bongers, E. H. Brilstra, C. W. Brown, H. T. Brüggenwirth, B. Callewaert, C. Chiang, K. Corning, H. Cox, E. Cuppen, B. B. Currall, T. Cushing, D. David, M. A. Deardorff, A. Dheedene, M. D'Hooghe, B. B. de Vries, D. L. Earl, H. L. Ferguson, H. Fisher, D. R. FitzPatrick, P. Gerrol, D. Giachino, J. T. Glessner, T. Gliem, M. Grady, B. H. Graham, C. Griffis, K. W. Gripp, A. L. Gropman, A. Hanson-Kahn, D. J. Harris, M. A. Hayden, R. Hill, R. Hochstenbach, J. D. Hoffman, R. J. Hopkin, M. W. Hubshman, A. M. Innes, M. Irons, M. Irving, J. C. Jacobsen, S. Janssens, T. Jewett, J. P. Johnson, M. C. Jongmans, S. G. Kahler, D. A. Koolen, J. Korzelius, P. M. Kroisel, Y. Lacassie, W. Lawless, E. Lemyre, K. Leppig, A. V. Levin, H. Li, H. Li, E. C. Liao, C. Lim, E. J. Lose, D. Lucente, M. J. Macera, P. Manavalan, G. Mandrile, C. L. Marcelis, L. Margolin, T. Mason, D. Masser-Frye, M. W. McClellan, C. J. Mendoza, B. Menten, S. Middelkamp, L. R. Mikami, E. Moe, S. Mohammed, T. Mononen, M. E. Mortenson, G. Moya, A. W. Nieuwint, Z. Ordulu, S. Parkash, S. P. Pauker, S. Pereira, D. Perrin, K. Phelan, R. E. Aguilar, P. J. Poddighe, G. Pregno, S. Raskin, L. Reis, W. Rhead, D. Rita, I. Renkens, F. Roelens, J. Ruliera, P. Rump, S. L. Schilit, R. Shaheen, R. Sparkes, E. Spiegel, B. Stevens, M. R. Stone, J. Tagoe, J. V. Thakuria, B. W. van Bon, J. van de Kamp, I. van Der Burgt, T. van Essen, C. M. van Ravenswaaij-Arts, M. J. van Roosmalen, S. Vergult, C. M. Volker-Touw, D. P. Warburton, M. J. Waterman, S. Wiley, A. Wilson, M. C. Yerena-de Vega, R. T. Zori, B. Levy, H. G. Brunner, N. de Leeuw, W. P. Kloosterman, E. C. Thorland, C. C. Morton, J. F. Gusella, and M. E. Talkowski. 2017. The genomic landscape of balanced cytogenetic abnormalities associated with human congenital anomalies. Nat Genet 49:36–45.

Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. Trends in Ecology & Evolution, 16(7), 351–358.

Sedlazeck, F. J., P. Rescheneder, M. Smolka, et al. 2018. Accurate detection of complex structural variations using single-molecule sequencing. Nature Methods 15:461–468.

Shang, H., D. L. Field, O. Paun, M. Rendón-Anaya, J. Hess, C. Vogl, J. Liu, P. K. Ingvarsson, C. Lexer, and T. Leroy. 2023. Drivers of genomic landscapes of differentiation across a Populus divergence gradient. Molecular Ecology 32:4348–4361.

Shen W, Le S, Li Y, Hu F (2016) SeqKit: A Cross-Platform and Ultrafast Toolkit for FASTA/Q File Manipulation. PLOS ONE 11(10): e0163962.

Shi, J., Z. Jia, J. Sun, et al. 2023. Structural variants involved in high-altitude adaptation detected using single-molecule long-read sequencing. Nat Commun 14:8282.

Simão F.A., Waterhouse R.M., Ioannidis P., Kriventseva E.V., Zdobnov E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.

Sobel, J. M., G. F. Chen, L. R. Watt, and D. W. Schemske. 2010. The biology of speciation. *Evolution* 64:295–315.

SRA Toolkit Development Team. 2024. SRA Toolkit. Available from: https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?view=software

Sudmant, P. H., T. Rausch, E. J. Gardner, R. E. Handsaker, A. Abyzov, J. Huddleston, Y. Zhang, et al. 2015. An integrated map of structural variation in 2,504 human genomes. *Nature* 526:75–81.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Research* 34:W609–W612.

Tham, C., Tirado-Magallanes, R., Goh, Y. *et al*. 2020. NanoVar: accurate characterization of patients' genomic structural variants using low-depth nanopore sequencing. *Genome Biol* **21**, 56.

The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).

Villar, D., C. Berthelot, S. Aldridge, et al. 2015. Enhancer evolution across 20 mammalian species. *Cell* 160(3):554–566.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, Kissinger JC, Paterson AH. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Research, 40(7), e49.

Wellenreuther M, Mérot C, Berdan E, Bernatchez L. Going beyond SNPs: The role of structural genomic variants in adaptive evolution and species diversification. Mol Ecol. 2019; 28: 1203–1209.

Wold, J. R., J. G. Guhlin, P. K. Dearden, A. W. Santure, and T. E. Steeves. 2023. The promise and challenges of characterizing genome-wide structural variants: A case study in a critically endangered parrot. Mol Ecol Resour.

Yang, J., D. F. Wang, J. H. Huang, et al. 2024. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. Genome Biology 25:148.

Zhang, L., S. Chaturvedi, C. C. Nice, L. K. Lucas, and Z. Gompert. 2023. Population genomic evidence of selection on structural variants in a natural hybrid

Zhang, Y., Yang, Z., He, Y. et al. Structural variation reshapes population gene expression and trait variation in 2,105 Brassica napus accessions. 2024. Nat Genet 56, 2538–2550 .

Zuellig, M. P., and A. L. Sweigart. 2018. Gene duplicates cause hybrid lethality between sympatric species of Mimulus. PLoS Genet 14:e1007130.

**Figure caption:**

**Figure 1.** General workflow of this study depicted on the right. The map shows the geographical range of the two species under investigation. The light grey area represents the range of white-tailed deer, while the grey area indicates the range of mule deer. The dark grey region highlights the overlap between the two species' habitats. Each point indicates the specific locations where samples were collected.

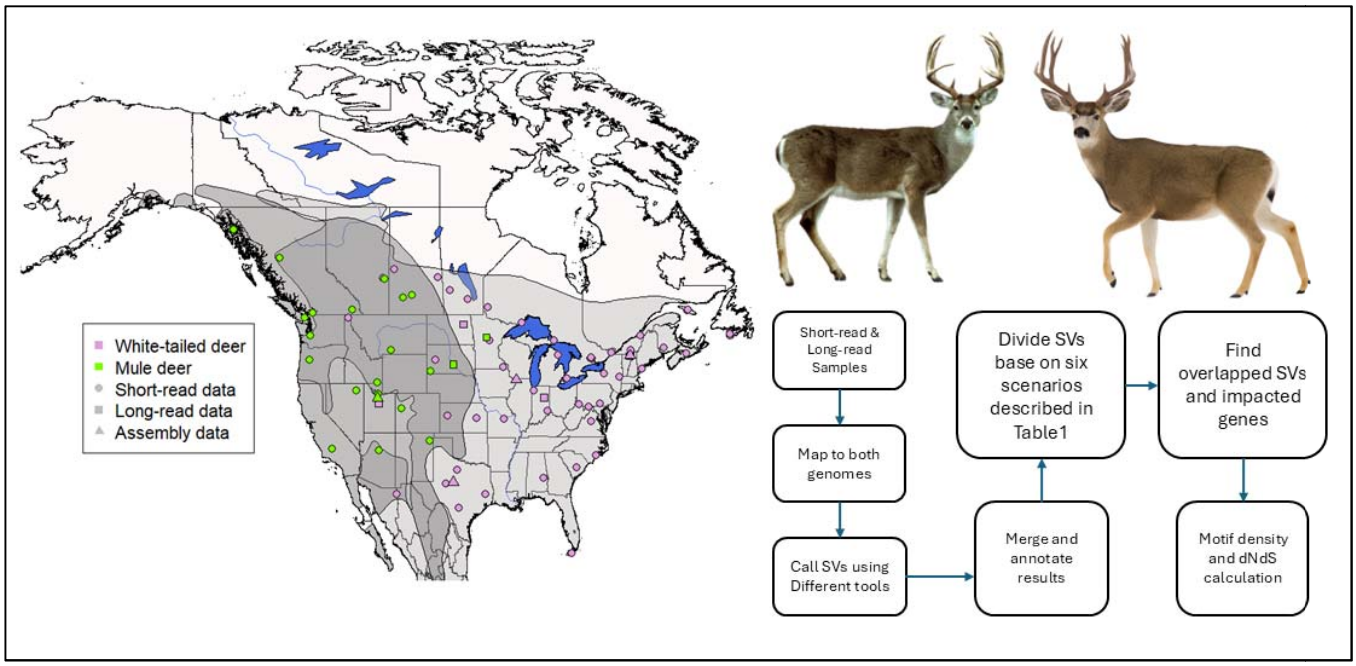



**Figure 2.** Macrosynteny comparison among white-tailed deer (*Odocoileus virginianus*), mule deer (*Odocoileus hemionus*), caribou (*Rangifer tarandus*), and domestic cattle (*Bos taurus*). Chromosome segments are connected by colored ribbons, showing conservation of chromosome structure and highlighting the close genomic relationship between white tailed and mule deer.

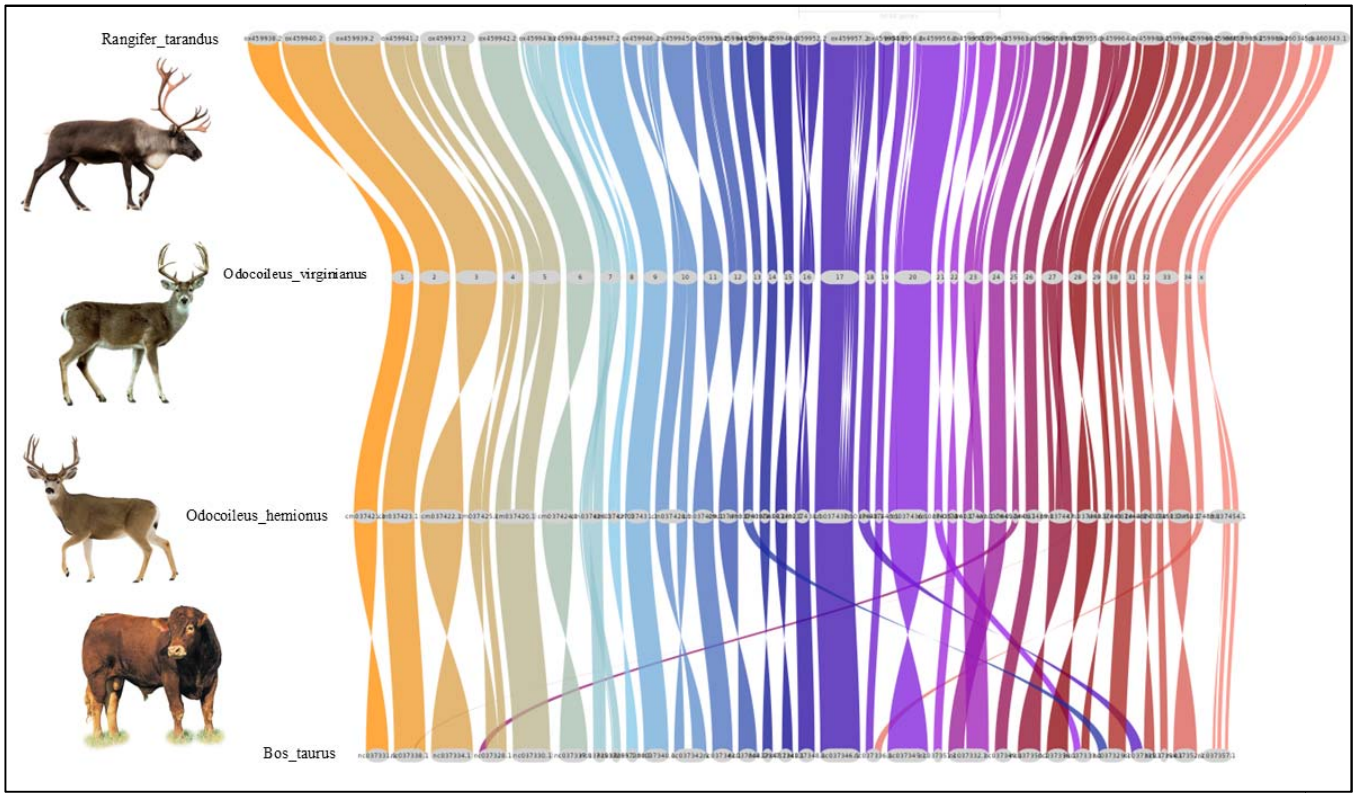



**Figure 3.** Comparison of the number of structural variants (SVs) detected across SV types in the entire short read dataset. Each panel represents a different mapping strategy, showing the counts of deletions (DEL), duplications (DUP), insertions (INS), inversions (INV), and translocations (TRA) on a log scale.

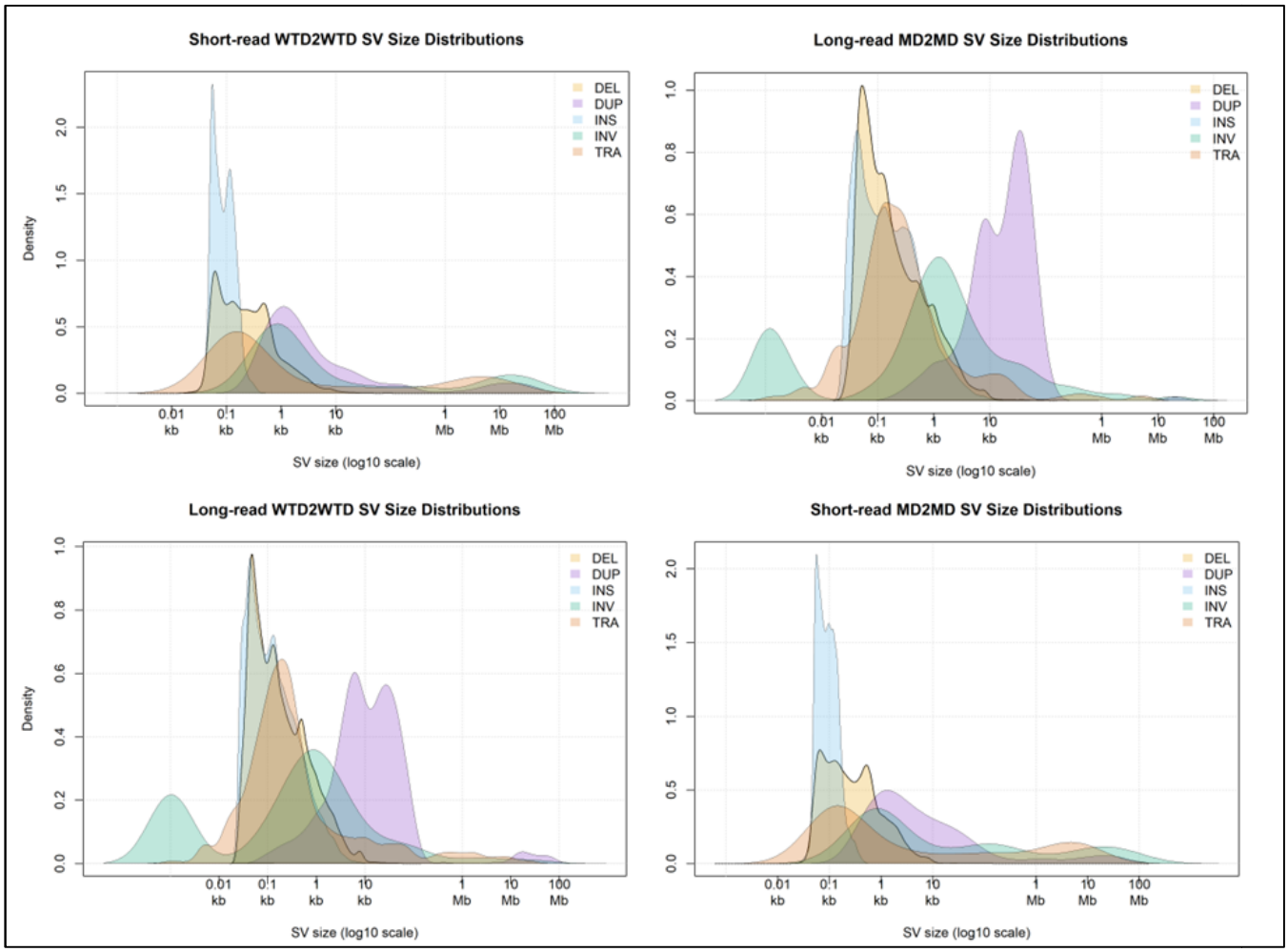

Short-read WTD2WTD SV Size Distributions
Long-read MD2MD SV Size Distributions
Long-read WTD2WTD SV Size Distributions
Short-read MD2MD SV Size Distributions
DEL
DUP
INS
INV
TRA
Density
SV size (log10 scale)
0.01 kb
0.1 kb
1 kb
10 kb
1 Mb
10 Mb
100 Mb

**Table captions:**

**Table 1.** Summary of the mapping scenarios used to identify species-specific structural variants (SVs) in both short-read and long-read datasets. WTD and MD refer to white-tailed deer and mule deer, respectively. The first four columns indicate the mapping strategy for each dataset—that is, which species' samples were aligned to which reference genome, with 1 meaning all samples have the SV. For example, "WTD2WTD" represents white-tailed deer samples mapped to the white-tailed deer reference genome. Last column indicates if the SVs derived from that strategy were considered species-specific (Y) or no (N).

| | WTD2WTD | WTD2MD | MD2WTD | MD2MD | SPECIES_SV |
|---|---|---|---|---|---|
| SCENARIO1 | 1 | 1 | 0 | 0 | Y* |
| SCENARIO2 | 0 | 0 | 1 | 1 | Y* |
| SCENARIO3 | 1 | 0 | 0 | 0 | N |
| SCENARIO4 | 0 | 1 | 0 | 0 | Y |
| SCENARIO5 | 0 | 0 | 1 | 0 | Y |
| SCENARIO6 | 0 | 0 | 0 | 1 | N |

***SV is absent or modified in both WTD and MD reference genomes.**

**Table 2.** Raw sequencing statistics for all long-read samples from white-tailed deer (WTD) and mule deer (MD). For each dataset, the total number of reads (No. reads), minimum read length (Min. length), average read length (Avg. Length), and maximum read length (Max length) are reported.

| Sample | Species | No. reads | Min. length | Avg. Length | Max length |
|---|---|---|---|---|---|
| bc2001 | WTD | 468,805 | 215 | 11,831.10 | 38,337 |
| bc2065 | MD | 1,820,410 | 51 | 11,292.10 | 52,827 |
| bc2066 | WTD | 1,144,838 | 76 | 13,229.60 | 47,420 |
| bc2069 | WTD | 1,735,556 | 64 | 12,179.30 | 41,157 |
| bc2070 | MD | 1,909,554 | 77 | 12,406.30 | 42,022 |
| SRR15720448 | MD | 15,003,446 | 50 | 19,657.10 | 256,018 |
| SRR6668252 | WTD | 89,770,713 | 11 | 5,093 | 217,085 |

**Table 3.** Species-specific structural variants (SVs) detected in short-read dataset across the scenarios defined in Table 1, using the criterion that each SV must be present in 100% of samples. For each scenario, the table reports: (i) the total number of SVs; (ii) counts of each SV class (DEL, DUP, INS, INV, TRA); (iii) the number of genes overlapping those SVs; and (iv) the distribution of gene impacts, categorized as exonic, intronic or gene-level. SV types are abbreviated as follows: DEL = deletion, DUP = duplication, INS = insertion, INV = inversion, TRA = translocation.

**SR - 100% of samples**

| | # SVs | SV Types | # Genes | Gene Impact |
|---|---|---|---|---|
| *Scenario1* | 11 | DEL:0 DUP:0 INS:3 INV:0 TRA:8 | 11 | 0\3\8 |
| *Scenario2* | 10 | DEL:1 DUP:0 INS:0 INV:1 TRA:8 | 9 | 0\0\9 |
| *Scenario4* | 4 | DEL:2 DUP:0 INS:1 INV:0 TRA:1 | 3 | 0\0\3 |
| *Scenario5* | 31 | DEL:13 DUP:6 INS:0 INV:2 TRA:10 | 17 | 4\1\12 |

**Table 4.** Species-specific structural variants (SVs) detected in the long-read dataset across the scenarios defined in Table 1, using the criterion that each SV must be present in 100% of samples. For each scenario, the table reports: (i) the total number of SVs; (ii) counts of each SV class (DEL, DUP, INS, INV, TRA); (iii) the number of genes overlapping those SVs; and (iv) the distribution of gene impacts, categorized as exonic, intronic, or gene-level. SV types are abbreviated as follows: DEL = deletion, DUP = duplication, INS = insertion, INV = inversion, TRA = translocation.

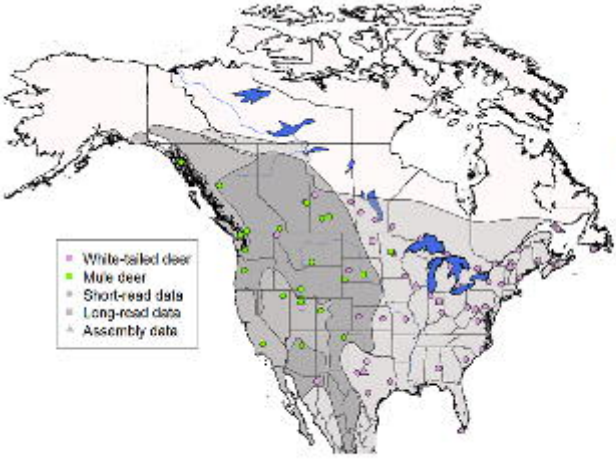
**LR - 100% of samples**

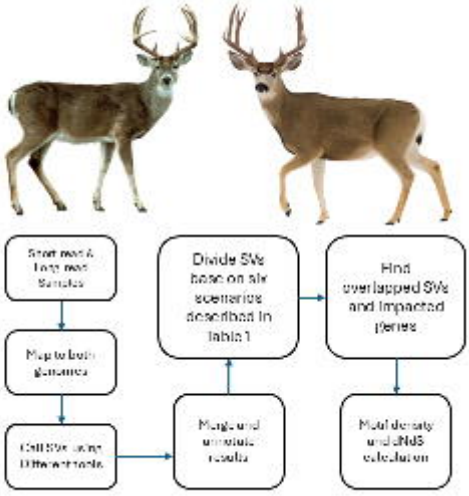
| | # SVs | SV Types | # Genes | Gene Impact |
|---|---|---|---|---|
| *Scenario1* | 19280 | DEL:8341 DUP:0 INS:10801 INV:138 TRA:0 | 4371 | 241\4127\3 |
| *Scenario2* | 14552 | DEL:7911 DUP:99 INS:4156 INV:217 TRA:2169 | 6220 | 426\3570\2224 |
| *Scenario4* | 23074 | DEL:14646 DUP:0 INS:8381 INV:47 TRA:0 | 2316 | 79\2050\187 |
| *Scenario5* | 66478 | DEL:37002 DUP:0 INS:29295 INV:181 TRA:0 | 12134 | 1307\10716\111 |

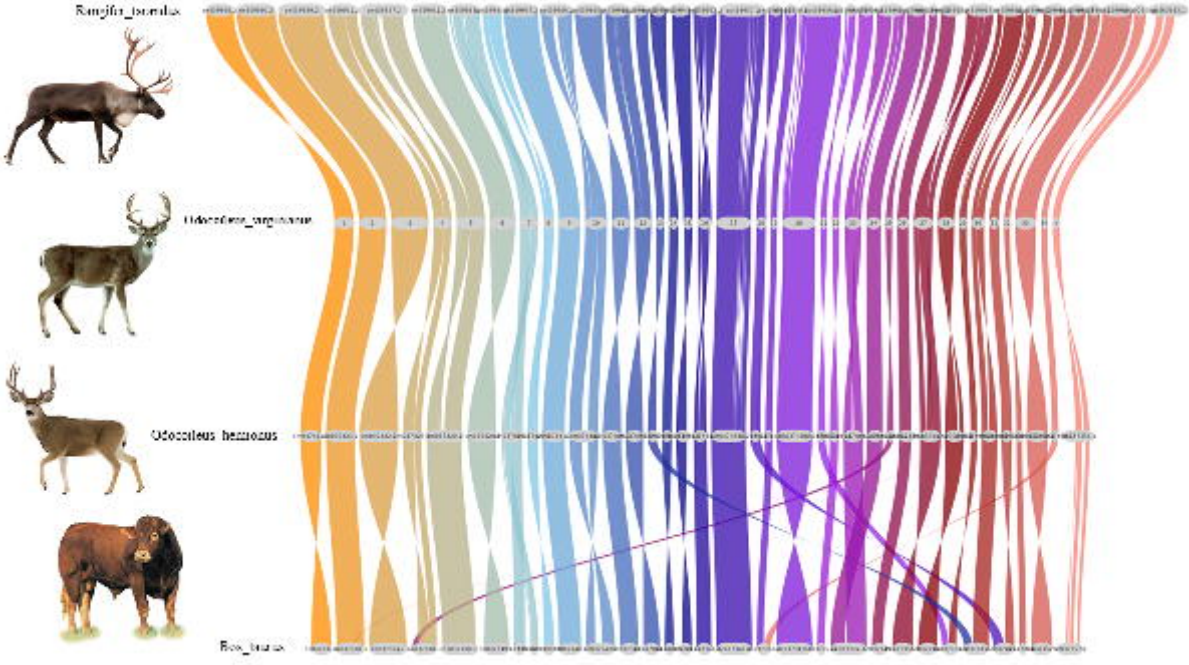
**Table 5.** Species-specific structural variants (SVs) detected in both long and short read datasets across the scenarios defined in Table 1. For each scenario, the table reports: (i) the total number of SVs; (ii) counts of each SV class (DEL, DUP, INS, INV, TRA); (iii) the number of genes overlapping those SVs; and (iv) the distribution of predicted gene impacts, categorized as exonic, intronic, or gene-level. SV types are abbreviated as follows: DEL = deletion, DUP = duplication, INS = insertion, INV = inversion, TRA = translocation.

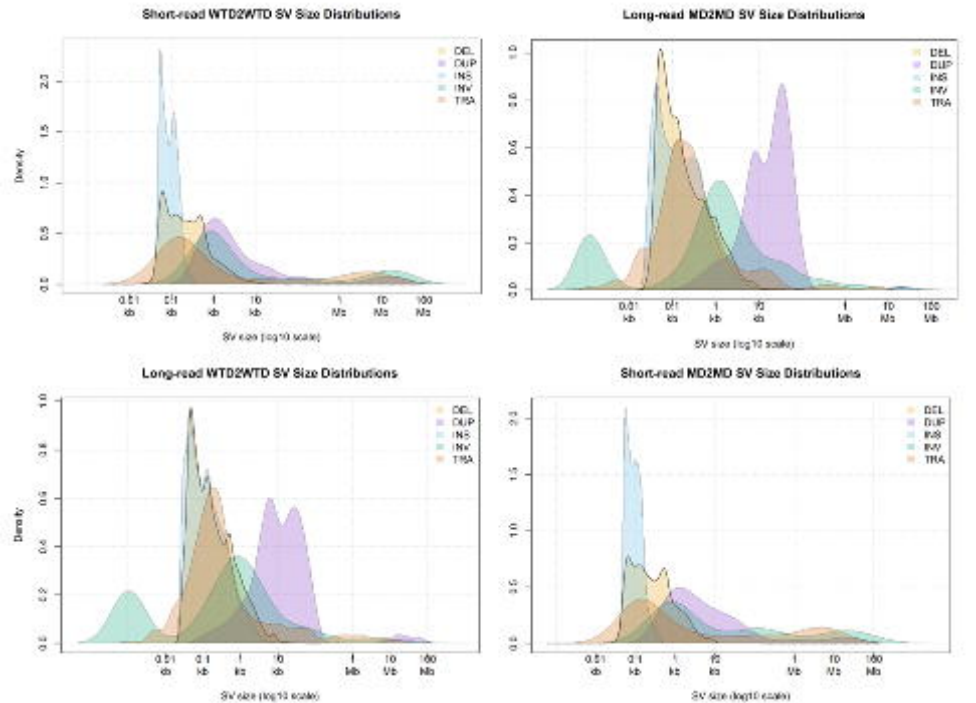
**Overlap structural variants**

| | # SVs | SV Types | # Genes | Gene Impact |
|---|---|---|---|---|
| *Scenario1* | 4 | INS:1 TRA:3 | 4 | 0\1\3 |
| *Scenario2* | 5 | TRA:5 | 5 | 0\0\5 |
| *Scenario4* | 3 | DEL:2 TRA:1 | 3 | 0\0\3 |
| *Scenario5* | 15 | DEL:3 DUP:2 INV:2 TRA:8 | 15 | 4\1\10 |


White-tailed deer
Mule deer
Short-read data
Long-read data
Assembly data


Divide SVs
Find
overlapped SVs
and impacted
genes


Rangifer tarandus
Odocoileus hemionus


Short-read WTD2WTD SV Size Distributions
Long-read MD2MD SV Size Distributions
Long-read WTD2WTD SV Size Distributions
Short-read MD2MD SV Size Distributions
DEL
DUP
INS
INV
TRA
Density
SV size (log10 scale)